

# The Power of Noise: Redefining Retrieval for RAG Systems

Year: 2024 | Citations: 287 | Authors: Florin Cuconasu, Giovanni Trappolini, F. Siciliano, Simone Filice, Cesare Campagnano

---

## Abstract

Retrieval-Augmented Generation (RAG) has recently emerged as a method to extend beyond the pre-trained knowledge of Large Language Models by augmenting the original prompt with relevant passages or documents retrieved by an Information Retrieval (IR) system. RAG has become increasingly important for Generative AI solutions, especially in enterprise settings or in any domain in which knowledge is constantly refreshed and cannot be memorized in the LLM. We argue here that the retrieval component of RAG systems, be it dense or sparse, deserves increased attention from the research community, and accordingly, we conduct the first comprehensive and systematic examination of the retrieval strategy of RAG systems. We focus, in particular, on the type of passages IR systems within a RAG solution should retrieve. Our analysis considers multiple factors, such as the relevance of the passages included in the prompt context, their position, and their number. One counter-intuitive finding of this work is that the retriever's highest-scoring documents that are not directly relevant to the query (e.g., do not contain the answer) negatively impact the effectiveness of the LLM. Even more surprising, we discovered that adding random documents in the prompt improves the LLM accuracy by up to 35%. These results highlight the need to investigate the appropriate strategies when integrating retrieval with LLMs, thereby laying the groundwork for future research in this area.